**ANALISIS KOMPARASI PEMODELAN TOPIK TERJEMAH INDONESIA HADIST BUKHORI DENGAN BERTOPIC, LATEN SEMANTIC ANALYSIS (LSA) DAN LATENT DIRECHLET ALLOCATION (LDA)**

**Proposal Tesis**

**Diajukan untuk memenuhi syarat memperoleh gelar Magister Komputer pada Program Studi Teknik Informatika S-2**

****

OLEH :

**ASEP RIDWAN HIDAYAT**

**231012050036**

**PROGRAM STUDI TEKNIK INFORMATIKA S-2**

**PROGRAM PASCASARJANA**

**UNIVERSITAS PAMULANG**

**TANGERANG SELATAN**

**2024**

LEMBAR PERSETUJUAN PROPOSAL TESIS

**ANALISIS KOMPARASI PEMODELAN TOPIK TERJEMAH INDONESIA HADIST BUKHORI DENGAN BERTOPIC, LATEN SEMANTIC ANALYSIS (LSA) DAN LATENT DIRECHLET ALLOCATION (LDA)**

Telah disetujui untuk disidangkan pada Program Studi Teknik Informatika S-2

Program Pascasarjana Universitas Pamulang

Pada tanggal …………………

Oleh

Nama Asep Ridwan Hidayat

231012050036

Proposal Tesis ini telah disetujui oleh Pembimbing untuk ujian proposal:

|  |  |  |
| --- | --- | --- |
| Pembimbing I |  | Pembimbing II |
|  |  |  |
| Dr. Tukiyat, M.Si |  | Dr. Taswanda Taryo, M.Sc. |
| NIDN. 8898211019 |  | NIDN. |
|  |  |  |
| Disahkan:  Direktur Program Pascasarjana  Universitas Pamulang | | |
|  |  |  |
| Dr. Sajarwo Anggai, MT. | | |
| NIDN. xxxxxxx | | |

LEMBAR PENGESAHAN TESIS

**ANALISIS KOMPARASI PEMODELAN TOPIK TERJEMAH INDONESIA HADIST BUKHORI DENGAN BERTOPIC, LATEN SEMANTIC ANALYSIS (LSA) DAN LATENT DIRECHLET ALLOCATION (LDA)**

Telah dipertahankan di hadapan Dewan Penguji Program Program Studi Teknik Informatika S-2 Program Pascasarjana Universitas Pamulang

Pada tanggal …………………

Oleh  
Nama Mahasiswa

2210120000015

|  |  |  |
| --- | --- | --- |
| Penguji I |  | Penguji II |
| ……………………………………… |  | ……………………………………… |
| NIDN. ……………………………… |  | NIDN. ……………………………… |
|  |  |  |
| Pembimbing I |  | Pembimbing II |
| …………………………………… |  | ……………………………………… |
| NIDN. ……………………………… |  | NIDN. ……………………………… |
|  |  |  |
| Disahkan:  Direktur Program Pascasarjana  Universitas Pamulang | | |
|  |  |  |
| ……………………………………… | | |
| NIDN. ……………………………… | | |

LEMBAR PERNYATAAN TESIS

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah tesis ini dan disebutkan dalam daftar pustaka.

Tangerang Selatan, September 2024

Tanda tangan

(Materai Rp. 10.000)

**Asep Ridwan Hidayat**

NIM : 231012050036

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, karena atas berkat dan rahmat -Nya penulis dapat menyelesaikan tesis ini. Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, mulai dari masa perkuliahan hingga penyusunan tesis ini, sangatlah sulit bagi penulis untuk menyelesaikannya. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Universitas Pamulang dan Program Studi Teknik Informatika S-2 yang telah melayani proses akademik dan pembelajaran dengan baik, mulai dari pendaftaran mahasiswa baru, pelaksanaan perkuliahan, hingga penyusunan tugas akhir.
2. Rektor Universitas Pamulang yang telah mengizinkan penulis untuk menempuh studi program S-2.
3. Dr. Sajarwo, Anggai, MT, selaku Ketua Program Studi Teknik Informatika S-2 Universitas Pamulang.
4. Dr. Tukiyat, M.Si. dan Dr. Taswanda Taryo, M.Sc.selaku dosen pembimbing I dan II yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulisan dan penyusunan tesis ini.
5. Rekan-rekan kerja serta mahasiswa Program Studi Teknik Informatika S-2 Universitas Pamulang yang telah banyak mendukung penulis dalam menyelesaikan tesis ini.
6. Semua pihak yang terlibat dan tidak penulis sebutkan satu per satu.

Akhir kata, penulis berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga tesis ini membawa manfaat bagi pengembangan ilmu pengetahuan.

Penulis

Asep Ridwan Hidayat

PERNYATAAN PERSETUJUAN PUBLIKASI TESIS

Saya yang bertanda tangan di bawah ini:

Nama : Asep Ridwan Hidayat

NPM : 231012050036

Program Studi : Teknik Informatika S-2

Jenis Karya : Tesis

Sebagai sivitas akademik Universitas Pamulang, dengan ini saya menyetujui untuk memberikan kepada Universitas Pamulang Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

"ANALISIS KOMPARASI PEMODELAN TOPIK TERJEMAH INDONESIA HADIST BUKHORI DENGAN *BERTOPIC*, *LATEN SEMANTIC ANALYSIS* (LSA) DAN *LATENT DIRECHLET ALLOCATION* (LDA)

"

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Pamulang berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Tangerang Selatan

Pada tanggal: .....

Yang menyatakan,

(Asep Ridwan Hidayat)

ABSTRAK

Penelitian ini bertujuan untuk menganalisis dan membandingkan performa tiga metode pemodelan topik, yaitu *BerTopic*, *Latent Semantic Analysis* (LSA), dan *Latent Dirichlet Allocation* (LDA), dalam mengelompokkan dan mengidentifikasi topik-topik dari terjemahan Indonesia Hadis Bukhari. Pemodelan topik merupakan salah satu pendekatan penting dalam pemrosesan bahasa alami yang memungkinkan pemahaman lebih mendalam terhadap konten teks dengan mengekstraksi tema-tema utama. *BerTopic* adalah metode yang memanfaatkan representasi teks berbasis transformer dan clustering hierarkis untuk menghasilkan topik yang lebih akurat dan interpretatif. LSA, sebagai teknik berbasis dekomposisi matriks, mengidentifikasi hubungan laten antar kata dan dokumen dengan mengurangi dimensi data. Sementara itu, LDA merupakan model probabilistik yang memodelkan dokumen sebagai campuran dari beberapa topik dan mengasumsikan distribusi kata pada topik tertentu. Dalam penelitian ini, ketiga metode tersebut diterapkan pada dataset terjemahan Hadis Bukhari, dan hasilnya dibandingkan berdasarkan beberapa metrik seperti koherensi topik, akurasi klasifikasi, dan kemudahan interpretasi. Hasil analisis menunjukkan bahwa setiap metode memiliki kelebihan dan kekurangan yang berbeda, di mana BerTopic unggul dalam hal presisi topik, LSA memberikan dimensi laten yang abstrak, sementara LDA menghasilkan topik yang lebih terstruktur. Kesimpulan dari penelitian ini memberikan wawasan tentang metode terbaik untuk pemodelan topik teks religius dengan bahasa yang kompleks.

Kata Kunci: Topik Model, *BerTopic*, *Latent Semantic Analysis*, *Latent Dirichlet Allocation*, *Hadist Bukhari*

*ABSTRACT*

This study aims to analyze and compare the performance of three topic modeling methods, namely BerTopic, Latent Semantic Analysis (LSA), and Latent Dirichlet Allocation (LDA), in clustering and identifying topics from the Indonesian translation of Hadith Bukhari. Topic modeling is one of the important approaches in natural language processing that allows a deeper understanding of text content by extracting key themes. BerTopic is a method that utilizes transformer-based text representation and hierarchical clustering to generate more accurate and interpretative topics. LSA, as a matrix decomposition-based technique, identifies latent relationships between words and documents by reducing the dimensionality of the data. Meanwhile, LDA is a probabilistic model that models documents as a mixture of several topics and assumes the distribution of words on a particular topic. In this study, the three methods are applied to the Hadith Bukhari translation dataset, and the results are compared based on several metrics such as topic coherence, classification accuracy, and ease of interpretation. The analysis results show that each method has different advantages and disadvantages, where BerTopic excels in topic precision, LSA provides abstract latent dimensions, and LDA produces more structured topics. The conclusion of this study provides insight into the best method for topic modeling of religious texts with complex language.

*Keywords: Topic Model, BerTopic, Latent Semantic Analysis, Latent Dirichlet Allocation, Hadith Bukhari*

DAFTAR ISI

[LEMBAR PERSETUJUAN PROPOSAL TESIS ii](#_Toc177735747)

[LEMBAR PENGESAHAN TESIS iii](#_Toc177735748)

[LEMBAR PERNYATAAN TESIS iv](#_Toc177735749)

[KATA PENGANTAR v](#_Toc177735750)

[PERNYATAAN PERSETUJUAN PUBLIKASI TESIS vi](#_Toc177735751)

[ABSTRAK (200 kata – 300 kata) vii](#_Toc177735752)

[*ABSTRACT* viii](#_Toc177735753)

[DAFTAR ISI ix](#_Toc177735754)

[DAFTAR TABEL xii](#_Toc177735755)

[DAFTAR GAMBAR xiv](#_Toc177735756)

[BAB I PENDAHULUAN 1](#_Toc177735757)

[1.1 Latar Belakang 1](#_Toc177735758)

[1.2 Permasalahan Penelitian 7](#_Toc177735759)

[1.2.1 Identifikasi Masalah 7](#_Toc177735760)

[1.2.2 Ruang Lingkup Masalah 8](#_Toc177735761)

[1.2.3 Rumusan Masalah 8](#_Toc177735762)

[1.3 Tujuan dan Manfaat Penelitian 8](#_Toc177735763)

[1.3.1 Tujuan Penelitian 8](#_Toc177735764)

[1.3.2 Manfaat Penelitian 9](#_Toc177735765)

[1.4 Sistematika Penulisan 10](#_Toc177735766)

[BAB II LANDASAN TEORI 12](#_Toc177735767)

[2.1 Tinjauan Pustaka 12](#_Toc177735768)

[((konsep yang ada ditulis SLR dan ada tabel). Sitasi minimal 30 – 40 paper, buku dam sumber ilmiah yang dapat dipercaya) 12](#_Toc177735769)

[2.2 Landasan Teori 23](#_Toc177735770)

[2.2.1 Analisis Sentimen 23](#_Toc177735771)

[2.2.2 Jasa Pengiriman 24](#_Toc177735772)

[2.2.3 JNE 25](#_Toc177735773)

[2.2.4 Alur Pengiriman JNE 27](#_Toc177735774)

[2.2.5 *Google Play* 29](#_Toc177735775)

[2.2.6 *Online Review* 30](#_Toc177735776)

[2.2.7 *Web Scraping* 31](#_Toc177735777)

[2.2.8 *Machine Learning* 31](#_Toc177735778)

[2.2.9 *Text Mining* 32](#_Toc177735779)

[2.2.10 *Text Preprocessing* 33](#_Toc177735780)

[2.2.11 *Clasification* 33](#_Toc177735781)

[2.2.12 *Data Mining* 35](#_Toc177735782)

[2.2.13 *Naïve Bayes Clasifier* 35](#_Toc177735783)

[2.2.14 *KNN* 37](#_Toc177735784)

[2.2.15 Fitur dan Pembobotan 38](#_Toc177735785)

[2.2.16 Asosiasi Kata 40](#_Toc177735786)

[2.2.17 *Diagram Fishbone* 40](#_Toc177735787)

[2.2.18 *Google Colab* 41](#_Toc177735788)

[2.3 Kerangka Pemikiran 43](#_Toc177735789)

[2.4 Hipotesis 44](#_Toc177735790)

[BAB III METODE PENELITIAN 45](#_Toc177735791)

[3.1 Analisis Kebutuhan 45](#_Toc177735792)

[3.1.1 Populasi dan Sampel 45](#_Toc177735793)

[3.1.2 Metode Pengumpulan Data 45](#_Toc177735794)

[3.1.3 Jenis dan Sumber Data 46](#_Toc177735795)

[3.1.4 Perancangan Penelitian 47](#_Toc177735796)

[3.1.5 Metode Analisis Data 49](#_Toc177735797)

[3.2 *Scraping Data* 49](#_Toc177735798)

[3.3 *Preprocessing* 50](#_Toc177735799)

[3.3.1 Model Akurasi Data *dengan Confusion Matrix* 50](#_Toc177735800)

[3.3.2 Akurasi (*Accuracy*) 51](#_Toc177735801)

[3.3.3 Presisi (*Precision*) 51](#_Toc177735802)

[3.3.4 *Recall* (Sensitivitas atau *True* *Positive* *Rate*) 51](#_Toc177735803)

[3.3.5 *F1 Score* 52](#_Toc177735804)

[DAFTAR PUSTAKA 53](#_Toc177735805)

DAFTAR TABEL

[**Tabel 1.1.** Data *Top Brand* Jasa Kurir Indonesia 3](#_Toc177735806)

DAFTAR GAMBAR

[Gambar 2.1. Tahapan Proses Pengiriman JNE 27](#_Toc177736102)

[Gambar 2.2. Kerangka Pemikiran 43](#_Toc177736103)

1. PENDAHULUAN
   1. Latar Belakang

Hadis merupakan salah satu sumber utama ajaran Islam setelah Al-Qur'an yang memberikan pedoman hidup bagi umat Muslim. Dalam konteks perkembangan teknologi informasi, pengolahan teks hadis secara digital semakin dibutuhkan untuk mendukung penelitian dan aplikasi keagamaan. Salah satu tantangan dalam analisis teks hadis, khususnya terjemahan hadis, adalah memahami struktur topik yang terkandung di dalamnya. model topik menawarkan pendekatan yang berbeda dalam melakukan pengelompokan topik dari kumpulan teks.

Pemodelan topik adalah teknik analisis teks yang dapat menemukan struktur topik tersembunyi dalam korpus. Teknik ini berguna untuk pelabelan otomatis pada himpunan data teks untuk klasifikasi atau untuk meningkatkan relevansi dari mesin pencari. Metode pemodelan topik awal dan masih populer hingga saat ini adalah Latent Dirichlet Allocation (LDA). Walaupun saat ini terdapat metode modern populer seperti BERTopic, namun LDA masih lebih populer digunakan dalam literatur Indonesia. Oleh karena itu diperlukan penelitian komparasi LDA dan BERTopic untuk mengetahui metode mana yang lebih efektif untuk teks berbahasa Indonesia dari segi nilai koherensi, nilai keberagaman, dan waktu pelatihan.

Penelitian ini bertujuan untuk membandingkan LDA dan BERTopic dalam memodelkan topik terjemah hadist bukhoru

seperti *Latent Dirichlet Allocation* (LDA) dan BERTopic menawarkan pendekatan yang berbeda dalam melakukan pengelompokan topik dari kumpulan teks. LDA menggunakan pendekatan statistik berdasarkan distribusi kata, sementara BERTopic menggabungkan model embedding berbasis BERT (*Bidirectional* *Encoder Representations from Transformers*) yang lebih kontekstual. Oleh karena itu, komparasi kedua metode ini pada terjemahan hadis Bukhari dalam bahasa Indonesia menjadi relevan untuk diinvestigasi guna menemukan pendekatan yang paling tepat dalam mengekstraksi dan mengelompokkan topik-topik dari teks hadis.

Teknologi informasi mengalami pertumbuhan yang cepat. Kemajuan teknologi mempermudah berbagai aspek kehidupan. Manusia harus terus menerus mengembangkan teknologi, mengingat kebutuhan terkait data dan informasi terus meningkat setiap tahunnya (Piliang, 2012). Tuntutan ini diperlukan untuk mendukung proses pengolahan data dan informasi agar menjadi lebih efektif dan efisien. Pertumbuhan teknologi yang pesat membawa peradaban manusia ke era digital. Era digital membawa dampak positif, seperti perkembangan teknologi internet yang penggunaannya terus meningkat setiap tahunnya (Prasetyo & Trisyanti, 2018). Perkembangan yang pesat ini juga mengalami transformasi yang berkelanjutan. Pertumbuhan bisnis tersebut menciptakan perubahan dalam pola dan strategi yang dijalani oleh pesaing untuk mempertahankan bisnis mereka. Oleh karena itu, setiap pelaku bisnis perlu memahami dinamika di sekitarnya dan fokus pada kemampuan untuk memengaruhi keputusan pembeli agar dapat mencapai kesuksesan dalam menjalankan usahanya.

Salah satu cara untuk mencapai keberhasilan dalam bisnis adalah melalui kegiatan pemasaran. Pemasaran berkaitan dengan mengidentifikasi dan memenuhi kebutuhan masyarakat, dengan definisi pemasaran yang paling sederhana adalah memenuhi kebutuhan pelanggan secara menguntungkan. Iklim persaingan perusahaan semakin ketat dan intensif, *term*asuk dalam penyediaan jasa logistik atau pengiriman kilat bagi masyarakat umum. Hal ini bertujuan untuk memberikan kemudahan dan kepuasan kepada masyarakat sebagai konsumen yang sangat penting dalam kehidupan, terutama di era globalisasi saat ini. Pada masa ini, masyarakat di dunia bisnis sibuk dengan berbagai kegiatan sehari-hari, sehingga muncul tuntutan untuk mengikuti perkembangan di mana segala sesuatu dilakukan dengan cepat dan praktis.

Tuntutan ini berlaku baik pada saat ini maupun di masa mendatang, terutama dalam memenuhi kebutuhan dan menjalankan usaha atau bisnis online. Faktor lain yang menjadi pertimbangan bagi pelanggan adalah kualitas layanan. Pelanggan akan memilih menggunakan jasa suatu perusahaan apabila perusahaan dapat memberikan pelayanan yang baik. Komponen yang membentuk kepuasan pelanggan ini khususnya relevan untuk perusahaan jasa. Kepuasan pelanggan akan tercapai jika mereka menerima pelayanan yang memuaskan atau sesuai dengan harapan mereka.

Dengan meningkatnya permintaan masyarakat terhadap layanan logistik belakangan ini, persaingan di antara perusahaan sejenis pun semakin meningkat. Oleh karena itu, setiap perusahaan harus mampu mengembangkan kemampuan, keterampilan, dan pengetahuannya untuk menghadapi persaingan dan memenuhi tuntutan konsumen yang semakin tinggi terhadap layanan ekspedisi. Tingginya minat dan kebutuhan masyarakat terhadap layanan ekspedisi barang membuat beberapa perusahaan logistik bersaing untuk memberikan kepuasan melalui pelayanan terbaik dan harga yang bersaing. Pemberian pelayanan yang unggul secara tidak langsung dapat memengaruhi keputusan konsumen dalam memilih penggunaan layanan logistik tersebut. Dari berbagai perusahaan jasa pengiriman yang dikenal oleh masyarakat, *JNE Express Across Nations* merupakan salah satu penyedia layanan pengiriman yang cukup terkenal. Ketika masyarakat menghadapi masalah terkait jasa pengiriman, mereka sering mengingat *JNE Express Across Nations* sebagai solusi pengiriman yang tersedia di Indonesia.

Teknologi turut ikut serta dalam mengubah gaya hidup manusia secara *fundamental*, salah satunya dalam aspek bisnis. Perusahaan memanfaatkan adanya teknologi informasi untuk menarik minat pelanggan. Peningkatan jumlah pengguna internet membuat berbagai perusahaan berlomba-lomba dalam memasarkan produknya, baik barang maupun jasa, melalui media internet. Berdasarkan survei yang dilakukan oleh APJII (2018) diketahui bahwa berjualan online *term*asuk ke dalam 10 besar alasan mengapa responden di Indonesia menggunakan internet. Sedangkan membeli secara online *term*asuk dalam 15 besar alasan utama penggunaan internet. Oleh karena itu terdapat banyak situs jual-beli secara online yang dikenal juga dengan sebutan *e-commerce*.

Pertumbuhan *e-commerce* ini tentunya ditopang oleh industri logistik, yaitu perusahaan ekspedisi. Kehadiran jasa ekspedisi memudahkan *e-commerce* mulai dari pergudangan hingga distribusi sehingga dapat lebih berfokus pada penjualan tanpa harus memikirkan pengiriman barang. Pada tahun 2017, kapitalisasi jasa ekspedisi menunjukkan bahwa salah satu industri jasa terbesar di Indonesia adalah pada bidang logistik (Chrisbiyanto, 2018).

Selain itu, Badan Pusat Statistik memaparkan bahwa pertumbuhan ekonomi di Indonesia sepanjang tahun 2017 mencapai 5,19% dengan sektor transportasi dan pergudangan menjadi salah satu yang paling tinggi. Perusahaan ekspedisi akan mengirimkan pesanan dari penjual agar sampai ke tangan konsumen. Dalam pengiriman barang ini pun segala sesuatunya terkontrol melalui teknologi informasi, mulai dari penyerahan barang ke petugas ekspedisi, *update* status terkini lokasi barang, hingga barang akhirnya diterima oleh konsumen.

Perkembangan jasa logistik juga dibuktikan melalui banyaknya jumlah perusahaan yang bergerak di bidang jasa pengiriman barang. Salah satu perusahaan jasa pengiriman dan logistik yang terkenal dan paling banyak digunakan adalah PT. Tiki Jalur Nugraha Ekakurir (JNE). Berdasarkan data *Top Brand* Indonesia terkait dengan jasa kurir yang paling banyak digunakan pada tahun 2015 sampai dengan 2019 menunjukkan bahwa JNE selalu berada pada urutan teratas seperti yang dapat dilihat pada Tabel 1.1.

**Tabel 1.1.** Data *Top Brand* Jasa Kurir Indonesia

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***BRAND*** | **2015** | **2016** | **2017** | **2018** | **2019** |
| JNE | 43.5% | 47.6% | 49.4% | 45.0% | 26.4% |
| J&T | - | - | - | 13.9% | 20.3% |
| Tiki | 36.2% | 35.7% | 34.7% | 13.6% | 12.6% |
| Pos  Indonesia | 6.7% | 9.6% | 8.4% | 11.6% | 5.4% |
| DHL | 2.1% | 1.3% | 1.3% | 3.5% | 3.8% |

JNE menjalin kerjasama dengan berbagai *e-commerce* di Indonesia, salah satunya adalah Tokopedia dan Shopee. JNE dipilih oleh masyarakat dibanding perusahaan lainnya bukan tanpa alasan. JNE, yang telah berdiri selama 29 tahun, memiliki jaringan yang luas di seluruh kota di Indonesia, dengan titik layanan 1.500 lokasi dan jumlah karyawan lebih dari 40.000 orang. Sistem komunikasi JNE telah terhubung secara online dengan lebih dari 150 lokasi dengan sistem yang efektif dan efisien untuk mengetahui status dari barang yang dikirim.

JNE memiliki aplikasi bernama *My JNE* yang dirilis pada tanggal 27 Januari 2016 di *Google Play*. Hingga saat ini, aplikasi *My JNE* telah diunduh oleh lebih dari satu juta pengguna. Adapun *My JNE* telah mendapat ulasan sebanyak lebih dari 86.000 komentar yang mencakup bukan hanya mengenai aplikasi, tetapi juga pelayanan JNE secara keseluruhan. Ulasan ini mencakup keluhan yang bersifat negatif dan juga saran yang positif. Ulasan pengguna merupakan salah satu media yang efektif dan efisien untuk menemukan informasi terhadap citra dari suatu perusahaan.

Hal ini dikarenakan konsumen melihat perspektif rekomendasi *word-of-mouth* (opini) pengguna lain sebelum menggunakan suatu produk maupun jasa (Fanani, 2017). Adapun rating yang didapatkan oleh JNE di *Google Play* hingga Desember 2019 hanya sebesar 2,4 bintang dari total keseluruhan penilaian 5 bintang.

*Market Share* Sistem Operasi di Indonesia menunjukkan bahwa mayoritas penduduk Indonesia menggunakan Android, mencapai persentase sebesar 50,61%. Android menyediakan berbagai layanan, salah satunya adalah *Google Play*, *platform digital* milik *Google* yang berfungsi sebagai toko untuk memasarkan berbagai produk, *term*asuk aplikasi, permainan, dan buku. Akses ke *Google Play* dapat dilakukan melalui *Play Store* (aplikasi Android), situs *web*, dan *Google* *TV* (Karch, 2019).

*Google Play* menawarkan beragam fitur, *term*asuk kemampuan pengguna untuk memberikan ulasan terhadap aplikasi. Ulasan ini memberikan wawasan tentang faktor-faktor yang memengaruhi pengguna dan area yang perlu diperbaiki oleh perusahaan.

Ulasan tersebut dapat dianalisis menggunakan teknik *Text Mining*, sebuah metode pencarian informasi berdasarkan data teks. Salah satu aspek dari *Text Mining* adalah analisis sentimen, yang memungkinkan klasifikasi opini ke dalam kategori positif, negatif, atau netral. Dengan demikian, perusahaan dapat melakukan perbaikan berdasarkan temuan sentimen negatif yang ditemukan dalam ulasan. Berdasarkan uraian di atas, perlu dilakukan analisis lebih lanjut terkait ulasan pengguna JNE di situs *Google Play* untuk mengetahui opini pengguna. Dilakukan pengambilan data dengan teknik *scraping* karena efektif dan efisien baik dari segi biaya, waktu, maupun tenaga. Kemudian dilakukan pengklasifikasian ulasan dalam bentuk positif dan negatif menggunakan algoritma *Naïve Bayes* *Classifier* dan *KNN*.

Menurut Wati (2016), *Naïve Bayes* merupakan metode paling sederhana dari pengklasifikasian probabilitas, memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada database dengan big data. Selain itu, model *Naïve Bayes* juga memiliki tingkat kesalahan yang sangat minimum dibandingkan dengan algoritma klasifikasi lainnya (Liu, et al., 2016).

Selanjutnya, dilakukan analisis dengan menggunakan *Diagram Fishbone* pada sentimen negatif untuk mengidentifikasi faktor sebab-akibat. Hal ini bertujuan agar JNE dapat memahami dengan lebih baik aspek-aspek yang memengaruhi ulasan negatif dan mengambil langkah-langkah yang diperlukan untuk meningkatkan kualitas layanannya.

Sentimen Analisis pada media sosial *Twitter* sudah dilakukan oleh beberapa peneliti, di antaranya yaitu penelitian Ayu Kusuma Dewi dan Sulastri, yang melakukan analisis sentimen ulasan masyarakat menggunakan data *scraping* dari aplikasi SiCepat di *Google Play*. Analisis menggunakan sentimen analisis metode *Naïve Bayes*. Kesimpulan penelitian ini adalah proses menjalankan klasifikasi diperlukan dua kelas sentimen, yaitu sentimen positif dan sentimen negatif. Pengumpulan data dilakukan dengan teknik *scraping* menggunakan AppFollow di *Google Play*, dengan data yang didapatkan sebanyak 457 ulasan positif dan negatif. Selanjutnya, pembagian data dibagi dua yaitu data latih dan data uji, dengan pembagian data sebesar 80% : 20%. Penerapan dalam pengujian text mining adalah untuk mengetahui hasil probabilitas dalam pengklasifikasian *Naïve Bayes*. Algoritma *Naïve Bayes* adalah proses melakukan klasifikasi dengan menggunakan metode probabilitas. Dari hasil penelitian didapatkan hasil tingkat akurasi sebanyak 80%, ulasan negatif sebanyak 87%, dan ulasan positif sebanyak 57%.

Penelitian selanjutnya dilakukan oleh Nuraini Ika Pratiwi Kalingara, dkk. Penelitian ini melakukan sentimen analisis masyarakat di *Twitter* terhadap komentar di aplikasi JNE dan J&T Express. Metode yang digunakan adalah metode *Naïve Bayes*. Kesimpulan dari penelitian ini adalah data akan dibagi dua: 75% *Data Training* dan 25% *Data Testing*. Kemudian dilakukan evaluasi dengan *Confusion Matrix*. Untuk JNE didapatkan hasil akurasi 79%, *Precision* 80%, *Recall* 79%, dan *F1-Score* 79%. Kemudian untuk J&T Express didapatkan hasil akurasi 76%, *Precision* 76%, *Recall* 76%, dan *F1-Score* 76%. Pada penelitian ini, diterapkan *K-Fold Cross Validation* dan menghasilkan skor untuk ekspedisi JNE sebesar 76% dan untuk ekspedisi J&T Express sebesar 75%. Dari persentase yang diperoleh, model masuk ke dalam klasifikasi baik.

Penelitian lain yang mirip adalah penelitian Farras Prasetiawan, dkk, yang melakukan sentimen analisis masyarakat di *Twitter* terhadap layanan ekspedisi SiCepat dengan menggunakan bahasa pemrograman *Python* *Google* *Colab*. Kesimpulan dari penelitian ini adalah proporsi sentimen SiCepat adalah sentimen positif sebesar 39% serta sentimen negatif sebesar 61%. Berdasarkan dimensi kualitas layanan SiCepat, jumlah proporsi yang paling banyak dibicarakan adalah dimensi kualitas layanan *reliability, assurance, responsiveness, empathy,* dan *tangibles*. Penelitian lain dilakukan oleh Nenden Legiawati, dkk, menggunakan metode klasifikasi data yaitu algoritma *Naïve Bayes* dengan menghitung nilai dari data *tweet* jasa ekspedisi Anteraja. Kesimpulan dari penelitian ini adalah hasil *Accuracy* 87,77%, *Precision* 76,67%, *Recall* 52,27%. JNE dengan *Accuracy* 81,48%, *Precision* 71,43%, dan *Recall* 62,50%. JNT dengan *Accuracy* 91,46%, *Precision* 48,15%, *Recall* 86,67%. Shopee Express dengan *Accuracy* 92,68%, *Precision* 9,09%, *Recall* 16,67% dan SiCepat dengan *Accuracy* 91,50%, *Precision* 100,00%, dan *Recall* 38,10%. ***Particle Swarm Optimization*** *(PSO)* berfungsi untuk meningkatkan nilai hasil dari klasifikasi *Naïve Bayes* dengan hasil dari Anteraja *Accuracy* 91,70%, *Precision* 82,05%, *Recall* 72,73%. JNE *Accuracy* 93,83%, *Precision* 88,00%, *Recall* 91,67%. JNT *Accuracy* 92,18%, *Precision* 70,97%, *Recall* 81,48%. Shopee Express *Accuracy* 94%, *Precision* 20,00%, *Recall* 33,33%, dan SiCepat *Accuracy* 95,42%, *Precision* 93,75%, *Recall* 71,43%. Dari hasil penelitian *Naïve Bayes* dan ***Particle Swarm Optimization*** *(PSO)* dapat dibandingkan bahwa ***Particle Swarm Optimization*** *(PSO)* terbukti mampu meningkatkan nilai dari *Naïve Bayes*.

Perbedaan penelitian yang sudah dilakukan sebelumnya dengan penelitian ini adalah pada penelitian ini dilakukan sentimen analisis opini pengguna terhadap jasa pengiriman JNE dan menggunakan 2 metode yaitu *Naïve Bayes* dan *KNN*. Pengambilan data juga dilakukan dengan cara *scraping* pada aplikasi JNE di *Google Play*, secara keseluruhan berbeda dengan penelitian lainnya. Perbedaan penelitian sebelumnya dengan penelitian ini adalah dalam mengambil *Dataset* dari media sosial *Twitter* seperti semua penelitian yang sudah dijelaskan. Dan penelitian yang lain hanya menggunakan 1 metode yaitu metode *Naïve Bayes* saja dan jasa ekspedisinya berbeda.

Penelitian ini diharapkan dapat menyajikan klasifikasi analisis sentimen yang akurat dan relevan, sehingga diperoleh informasi yang berharga terkait langkah-langkah yang dapat diambil oleh JNE untuk meningkatkan kualitas layanannya dan mengembangkan perusahaannya. Oleh karena itu, penulis mengangkat tema penelitian dalam bentuk tesis dengan judul **“ANALISIS SENTIMEN OPINI PENGGUNA JASA PENGIRIMAN JNE MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER* DAN *K-NEAREST NEIGHBORS*.”**

* 1. Permasalahan Penelitian

Berdasarkan latar belakang masalah diperoleh rumusan masalah penelitian yang akan diteliti sesuai dengan identifikasi dan dibatasi dalam suatu ruang lingkup permasalahan.

* + 1. Identifikasi Masalah

Berdasarkan paparan dari latar belakang penelitian di atas, dapat diidentifikasi masalah sebagai berikut :

1. Belum ada panduan yang cukup jelas mengenai cara menerapkan metode *Naïve Bayes* *Classifier* dan *K-Nearest Neighbors* pada data pengguna JNE berdasarkan ulasan di aplikasi *My JNE* di *web*site *Google Play*.
2. Performa metode *Naïve Bayes* *Classifier* dan *K-Nearest Neighbors* dalam mengklasifikasikan opini pengguna JNE menjadi opini positif dan negatif belum dapat diandalkan.
   * 1. Ruang Lingkup Masalah

Berdasarkan pada latar belakang masalah yang ada maka penelitian ini akan membatasi permasalahan yang ada, yaitu :

1. Data yang akan diklasifikasi hanya data ulasan pengguna JNE pada
2. *Google Play*
3. Ulasan pengguna mengenai JNE pada *web*site *Google Play* yang diambil hanya pada periode setelah aplikasi *My JNE* di *update* yaitu setelah 10 Maret 2024 sampai dengan 31 April 2024.
4. Menerapkan metode klasifikasi *Naïve Bayes* dan *KNN* dalam menginterpretasikan sentiment pada data yang di ambil.
5. Menganalisis sentimen positif, negatif, atau netral yang muncul dari respons pengguna JNE pada *Google* play.
   * 1. Rumusan Masalah

Berdasarkan identifikasi masalah di atas, maka dapat dibuat rumusan masalah sebagai berikut :

1. Bagaimana mengimplementasikan metode *Naïve Bayes* *Classifier* dan K- Nearest Neighbor pada data pengguna JNE berdasarkan ulasan pada aplikasi *My JNE* di *web*site *Google Play*?
2. Bagaimana performa metode *Naïve Bayes* *Classifier* dan *K-Nearest Neighbors* dalam mengklasifikasikan opini pengguna JNE menjadi opini positif, negatif dan netral?
   1. Tujuan dan Manfaat Penelitian
      1. Tujuan Penelitian

Tujuan penelitian ini adalah untuk memberikan jawaban atas identifikasi masalah yang telah disebutkan diatas, yaitu :

1. Untuk mengimplementasikan metode Naïve Bayes Classifier dan K-Nearest Neighbor (KNN) untuk mengklasifikasikan data ulasan pengguna pada aplikasi My JNE di Google Play Store.
2. Untuk mengevaluasi dan membandingkan performa metode Naïve Bayes Classifier dan K-Nearest Neighbor dalam mengklasifikasikan opini pengguna JNE menjadi kategori positif, negatif, dan netral.
   * 1. Manfaat Penelitian
3. Manfaat bagi Penulis:

* Pengembangan Keahlian:

Penelitian ini memberikan kesempatan bagi penulis untuk mengembangkan keterampilan analisis data dan pemahaman mendalam mengenai metode klasifikasi *Machine Learning*, seperti *Naïve Bayes* dan *K-Nearest Neighbors* (*KNN*).

* Peningkatan Kompetensi Akademik:

Melalui penelitian ini, penulis dapat memperkuat kompetensi akademik dan kemampuan dalam menyusun karya ilmiah yang berkualitas, yang akan berguna dalam karir akademik dan profesional di masa depan.

* Kontribusi terhadap Pengetahuan:

Penelitian ini memungkinkan penulis untuk berkontribusi pada literatur ilmiah, khususnya dalam bidang analisis sentimen dan penerapan *Machine Learning* pada data ulasan aplikasi.

1. Manfaat bagi Pihak Terkait:

* Peningkatan Kualitas Layanan JNE:

Hasil penelitian ini dapat digunakan oleh JNE untuk mengidentifikasi faktor-faktor yang menyebabkan ulasan negatif, sehingga perusahaan dapat mengambil langkah-langkah perbaikan untuk meningkatkan kualitas layanan dan kepuasan pelanggan.

* Pengambilan Keputusan Berdasarkan Data:

Pihak manajemen JNE dapat memanfaatkan hasil analisis sentimen ini sebagai dasar dalam pengambilan keputusan strategis untuk meningkatkan citra dan performa perusahaan di mata konsumen.

* Model untuk Analisis Sentimen:

Penelitian ini dapat dijadikan acuan oleh perusahaan lain dalam industri logistik atau bidang layanan lainnya untuk melakukan analisis sentimen terhadap opini pengguna dengan menggunakan metode serupa.

1. Manfaat bagi Kampus:

* Peningkatan Reputasi Akademik:

Penelitian ini akan berkontribusi pada peningkatan reputasi kampus dalam bidang riset, khususnya di bidang teknologi informasi dan analisis data, melalui publikasi hasil penelitian yang berkualitas.

* Sumber Referensi bagi Mahasiswa:

Hasil penelitian ini dapat menjadi sumber referensi bagi mahasiswa lain yang tertarik untuk melakukan penelitian dalam bidang yang sama, sehingga memperkaya literatur dan penelitian yang ada di kampus.

* Peningkatan Kerjasama dengan Industri:

Dengan adanya penelitian yang relevan dan bermanfaat bagi industri, kampus dapat meningkatkan kerjasama dengan berbagai pihak eksternal, *term*asuk perusahaan dan institusi lainnya, dalam bidang penelitian dan pengembangan teknologi.

* 1. Sistematika Penulisan

Penulisan tesis ini disusun secara sistematis untuk memberikan gambaran yang jelas mengenai seluruh tahapan penelitian yang dilakukan. Sistematika penulisan ini terdiri dari lima bab utama, yang masing-masing diuraikan sebagai berikut:

**BAB I - Pendahuluan**

Bab ini berisi latar belakang penelitian, perumusan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan. Pendahuluan memberikan konteks yang jelas mengenai pentingnya penelitian ini dan masalah-masalah yang ingin dipecahkan.

**BAB II - Landasan Teori dan Kerangka Pemikiran**

Berisi tentang landasan teori *Data Mining* dengan algoritma *Naïve Bayes* yang akan digunakan pada Analisis sentimen Jasa Pengiriman JNE

**BAB III - Metode Penelitian**

Bab ini dipaparkan tentang metode yang peneliti pakai pengumpulan data yang akan dilakukan pada penelitian ini.

**BAB IV - Hasil dan Pembahasan**

Memaparkan hasil percobaan sesuai dengan metodologi yang dirancang pada bab sebelumnya.

**BAB V - Kesimpulan dan Saran**

Bab ini menyimpulkan hasil penelitian dan menjawab pertanyaan penelitian yang diajukan. Kesimpulan yang diambil berdasarkan temuan penelitian disampaikan secara ringkas dan jelas. Selain itu, bab ini juga memberikan saran untuk penelitian lanjutan serta rekomendasi praktis bagi pihak-pihak yang berkepentingan.

1. LANDASAN TEORI
   1. Tinjauan Pustaka

((konsep yang ada ditulis SLR dan ada tabel). Sitasi minimal 30 – 40 paper, buku dam sumber ilmiah yang dapat dipercaya)

Penelitian mengenai analisis sentimen telah berkembang pesat, terutama dalam konteks penggunaan algoritma machine learning seperti Naïve Bayes, K-Nearest Neighbors (KNN), dan Support Vector Machines (SVM). Berbagai studi telah dilakukan untuk menganalisis sentimen dari berbagai data teks, termasuk ulasan pengguna di platform media sosial, ulasan aplikasi, dan opini publik di media online. Metode dan hasil dari penelitian ini menunjukkan bagaimana berbagai algoritma digunakan untuk mengklasifikasikan data menjadi sentimen positif, negatif, atau netral, serta memberikan saran untuk penelitian lebih lanjut di bidang ini. Penelitian sebelumnya yang dijadikan referensi mencakup berbagai metode dan pendekatan dalam analisis sentimen, khususnya yang menggunakan data dari media sosial seperti *Twitter*. Penelitian-penelitian tersebut memberikan dasar teori dan metodologi yang kuat untuk penelitian ini, serta menginformasikan tentang teknik-teknik *Preprocesing* data, implementasi algoritma *Machine Learning*, dan evaluasi kinerja *Classifiers*.

Penelitian tentang *"Analisis Sentimen Pengguna Twitter terhadap Dompet Elektronik dengan Metode Lexicon Based dan K-Nearest Neighbors"* mengkaji sentimen pengguna terhadap layanan dompet elektronik seperti OVO, Gopay, dan LinkAja. Hasil dari penelitian ini menunjukkan bahwa metode K-Nearest Neighbors menghasilkan akurasi tertinggi pada Gopay sebesar 94,05%, sementara OVO dan LinkAja memiliki akurasi yang lebih rendah. Kesimpulan dari penelitian ini adalah metode K-Nearest Neighbors cukup efektif dalam mengklasifikasikan sentimen pengguna terhadap layanan dompet elektronik. Peneliti menyarankan agar penelitian lebih lanjut dilakukan dengan memperluas dataset untuk meningkatkan akurasi dan validitas model (Salim, 2020).

**Ricardo Cuatanto dan Rudi Sutomo (2020)** dalam penelitian yang berjudul *"Comparative Analysis Between Naïve Bayes Algorithm and Decision Tree Loss Rate from Fire Disaster Data in DKI Jakarta Province"* membandingkan kinerja dua algoritma, yaitu Naïve Bayes dan Decision Tree, dalam mengklasifikasikan tingkat keparahan kerugian akibat kebakaran di Jakarta. Penelitian ini menemukan bahwa algoritma Decision Tree lebih unggul dengan akurasi sebesar 78%, dibandingkan dengan Naïve Bayes yang hanya mencapai 75%. Hasil ini menunjukkan bahwa Decision Tree lebih efektif dalam mengelompokkan data terkait bencana kebakaran. Saran dari penelitian ini adalah untuk meningkatkan kesiapsiagaan dan mitigasi kebakaran di kawasan perkotaan dengan memanfaatkan hasil klasifikasi ini (Cuatanto & Sutomo, 2020).

Dan seterusanya dibuat minimal 10 peneliti terdahulu yang dijadikan sebagai acuan utama dalam penelitian ini,,,,,,

Lanjutkan ,,,,,,

Dari beberapa penelitian terdahulu yang dijadikan acuan pustaka, berikut ini diberikan reviu paper seperti pada tabel tabel 2.1 yang menyajikan ringkasan dari beberapa penelitian relevan atau penelitian sebelumnya.

**Tabel 2.1.** Tinjauan Pustaka

| **No** | **Penulis** | **Tahun** | **Judul Penelitian** | **Model** | ***Accuracy*** |
| --- | --- | --- | --- | --- | --- |
| 1 | Ss. Salim | 2020 | Analisis Sentimen pengguna *Twitter* terhadap Dompet eletronik dengan Metode *Lexicon Based* dan *K-Nearest Neighbors* | *Lexicon Based dan K-Nearest Neighbors* | Hasil akhir perhitungan metode *K-Nearest Neighbors* dengan *Confusion Matrix* untuk *tweet* OVO diperoleh nilai akurasi positif sebesar 86,91% dan nilai akurasi negatif 13,09%, *tweet* Gopay diperoleh nilai akurasi sebesar 94,05% dan nilai akurasi negatif 5,95%, serta *tweet* LinkAja diperoleh nilai akurasi sebesar 76,31% dan nilai akurasi *negative* 23,69%. |
| 2 | Ricardo Cuatanto, Rudi Sutomo | 2020 | *Comparative Analysis Between Naïve Bayes*  *Algorithm and Decision Tree Loss Rate from Fire Disaster Data in DKI Jakarta Province* | *Naïve Bayes* | penelitian ini membantu meningkatkan teknik mitigasi kebakaran. Keamanan dan kesiapan menghadapi bencana kebakaran di kawasan ini harus ditingkatkan. Akurasi kedua prediksi yang dilakukan algoritma *Naïve Bayes* adalah 75%. Sebaliknya, akurasi algoritma *Decision Tree* sebesar 78% sehingga menghasilkan kesimpulan bahwa pendekatan *Decision Tree* lebih membantu dalam mengkategorikan tingkat keparahan kerugian akibat bencana kebakaran. Kata Kunci Bencana Kebakaran Algoritma *Naïve Bayes* Algoritma *Decision Tree* Klasifikasi Tingkat Kerugian. |
| 3 | Debby Alita, Auliya Rahman Isnain | 2020 | Pendeteksian Sarkasme Pada Proses Analisis Sentimen Menggunakan *Random Forest* *Classifier* | *Random Forest* *Classifier* | Hasil penelitian ini didapatkan peningkatan nilai rata- rata akurasi sebesar 16,61 %, nilai presisi sebesar 5,45 %, nilai *Recall* sebesar 9,64% dan kenaikan nilai *F1-Score* sebesar 11,27% dengan jumlah data sebanyak 2.027 dengan rincian data dengan label positif berjumlah 1023, data dengan label negatif berjumlah 587 dan data dengan label netral berjumlah 462. Data sarkasme didapatkan dari *tweet* dengan label positif yang kemudian diberikan label sarkasme atau tidak sarkasme dan didapat hasil label dengan jumlah keseluruhan berlabel sarkasme |
| 4 | Harun & Putri Ananda | 2021 | Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan *Naïve Bayes* dan *Decision Tree* | *Naïve Bayes* | analisa sentimen opini masyarakat tentang vaksinasi COVID-19 yang telah dilakukan, cenderung ke tanggapan negatif dengan nilai akurasi 100.00% menggunakan algoritma NBC dan 50.39% menggunakan algoritma *Naïve Bayes*. |
| 5 | RD Septiana, AB Susanto, T Tukiyat | 2021 | Analisis Sentimen Vaksinasi Covid-19 Pada *Twitter* Menggunakan *Naïve Bayes* *Classifier* Dengan *Feature Selection Chi- Squared Statistic* Dan *Particle Swarm Optimization* | *Chi-Squared Statistic, NBC, PSO* | Hasil Pengujian algoritma *Naïve Bayes* *Classifier* menggunakan seleksi fitur *Chi Squared Statistic* berbasis *forward selection* memiliki nilai akurasi yang lebih tinggi dibandingkan menggunakan seleksi. |
| 6 | H Setiawan, E Utami, S Sudarmawan | 2021 | Analisis Sentimen *Twitter* Kuliah *Online* Pasca Covid-19 Menggunakan Algoritma *Support Vector Machine* dan *Naïve Bayes* | *Naïve Bayes, KNN* | Untuk analisis sentimen peneliti menerapkan algorima naïve *Naïve Bayes* dan *KNN* dengan hasil peforma yang didapat pada algoritma *Naïve Bayes* akurasi 81,20%, waktu 9,00 detik, *Recall* 79,60% dan presisi 79,40% sedangkan untuk algoritma *KNN* mendapatkan nilai akurasi 85%, waktu 31,60 detik, *Recall* 84% dan presisi 83,60%, hasil peforma tersebut diperoleh pada iterasi ke 1 untuk naïve *Naïve Bayes* dan iterasi ke 423 untuk algoritma *KNN*. |
| 7 | S Samsir, A Ambiyar, U Verawardina, F Edi, R Watrianthos | 2021 | Analisis Sentimen Pembelajaran Daring Pada *Twitter* di Masa Pandemi COVID-19 Menggunakan Metode *Naïve Bayes* | *Naïve Bayes* | Menunjukkan bahwa pembelajaran daringmemiliki 30% sentimen positif, 69% sentimen negatif, dan 1% netral pada periode tersebut. Tingginya sentimen negatif dihasilkan karena ketidakpuasaan masyarakat terhadap pembelajaran daring. Beberapa twit menunjukkan kekecewaan dengan kata „stres‟ dan „malas‟ merupakan kata yang memiliki frekuensi tinggi dalam percakapan. |
| 8 | Wijaya & Dwiasnati | 2020 | Implementasi *Data Mining* dengan Algoritma *Naïve Bayes* pada Penjualan Obat | *Naïve Bayes* | Penelitian ini menggunakan *tools* *Rapidminner* versi 8 sebagai media untuk mengujidata yang akan diolah untuk mendapatkan hasil *Accuracy* dan nilai *ROC*. Nilai *Accuracy* tersebut menunjukkan di nilai 88.00%. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

* 1. Landasan Teori
     1. Analisis Sentimen

Analisis Sentimen atau opinion mining merupakan studi komputasi mengenai pendapat, perilaku, dan emosi seseorang terhadap entitas yang mana entitas tersebut dapat mewakili individu, kejadian atau topik (Indrayuni, 2018). Melalui analisis sentimen dapat diketahui opini publik terkait suatu hal baik berupa produk, jasa, aplikasi, film, musik, dan lain sebagainya melalui *review* pengguna pada berbagai situs. Analisis sentimen digunakan untuk melihat kecenderungan apakah pendapat atau opini pada suatu masalah cenderung *term*asuk dalam opini positif atau opini negatif (Kurniawan & Susanto, 2019).

Dlam konteks ini tuliskan teori-teori yang akan digunakan dan mendukung untuk penelitian saudara

Misal Teori Data *Mining*

Data mining adalah proses untuk mengekstrak informasi atau pengetahuan yang berguna dari kumpulan data yang besar. Secara umum, *Data Mining* melibatkan penerapan teknik-teknik statistik, matematika, dan *Machine Learning* untuk mengidentifikasi pola tersembunyi, tren, atau hubungan dalam data yang dapat digunakan untuk membuat keputusan yang lebih baik. Data mining berfungsi sebagai alat penting dalam berbagai bidang, *term*asuk bisnis, keuangan, pemasaran, kesehatan, dan penelitian ilmiah, untuk menganalisis data yang besar dan kompleks serta menemukan wawasan yang sebelumnya tidak terlihat.

Proses *Data Mining* terdiri dari beberapa tahap, mulai dari pengumpulan data, pembersihan dan persiapan data, hingga penerapan algoritma untuk menemukan pola atau model dalam data. Teknik-teknik yang digunakan dalam *Data Mining* antara lain *clustering*, *class*ification, *association rule mining*, dan *anomaly detection*. *Clustering* adalah proses pengelompokan data berdasarkan kesamaan, sedangkan *classification* melibatkan penentuan label atau kategori untuk data baru berdasarkan model yang sudah dilatih. *Association rule mining* digunakan untuk menemukan hubungan antara variabel dalam *Dataset*, dan *anomaly detection* bertujuan untuk mengidentifikasi data yang menyimpang dari pola umum.

Data mining menjadi semakin relevan seiring dengan pertumbuhan data yang cepat dalam era digital saat ini. Dengan kemampuannya untuk mengolah dan menganalisis data dalam jumlah besar, *Data Mining* memungkinkan organisasi dan individu untuk membuat keputusan berbasis data yang lebih informatif dan akurat. Di bidang penelitian, *Data Mining* digunakan untuk menemukan pola atau hubungan dalam data yang dapat memberikan wawasan baru atau mengonfirmasi hipotesis yang ada. Dengan demikian, *Data Mining* tidak hanya membantu dalam pengambilan keputusan, tetapi juga mendorong inovasi dan penemuan baru dalam berbagai disiplin ilmu.

* + 1. *Naïve Bayes Clasifier*

*Naïve Bayes* *Classifier* merupakan teknik klasifikasi berdasarkan Teorema *Bayes* dengan asumsi independensi di antara para prediktor. *Naïve Bayes* *Classifier* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema *Bayes*. *Naïve Bayes* merupakan metode paling sederhana dari pengklasifikasian probabilitas, memiliki tingkat akurasi yang sangat tinggi ketika diaplikasikan pada database dengan big data (Wati, 2016).Menurut Rennie et al.(2003) Menyatakan bahwa *Naïve Bayes* adalah salah satu model yang paling efisien dan tangguh dalam kalasifikasi text.yang di tekankan bahwa meskipun asumsi independensi yang mendasari model ini jarang terpenuhi,perporma yang dihasilkan sering memuaskan dalam aplikasi praktis.

Selain itu model naïve *Naïve Bayes* juga memiliki tingkat kesalahan yang sangat minimum dibandingkan dengan algoritma klasifikasi lainnya (Liu etal., 2016).

Algoritma naïve *Naïve Bayes* hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variabel independent, maka hanya variansi dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians. Rumus *Bayes* dalam (Muslehatindkk, 2017) secara umum adalah sebagai berikut:

(1)

dengan :

X = Data dengan *class* yang belum diketahui

H = Hipotesis data X merupakan suatu *class* spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (*prior prob*) P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Adapun aturan *Bayes* adalah sebagai berikut :

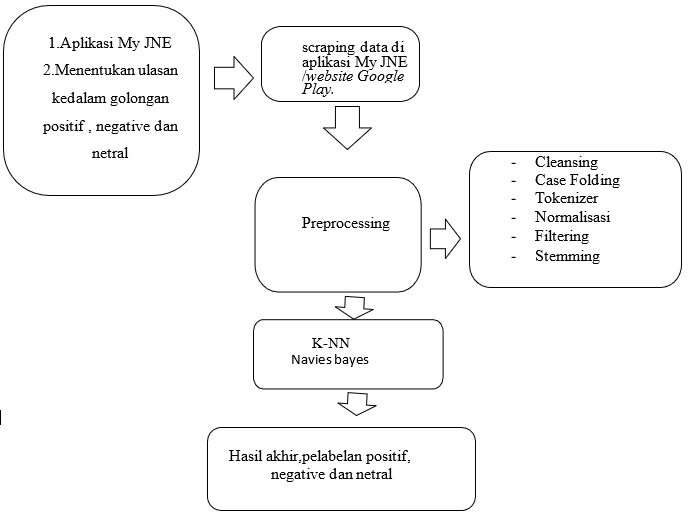
Jika P(h1|x) < P(h2|x), maka x diklasifikasikan sebagai h2. Pernyataan P(h1|x) mengindikasikan probabilitas hipotesis h1 berdasarkan kondisi x terjadi, begitu pula dengan h2. Sehingga didapat klasifikasi dari x sesuai dengan probabilitas terbesar diantara probabilitas x terhadap semua kelas.

Algoritma *Naïve Bayes* mengasumsikan bahwa fitur-fitur dalam data bersifat independen secara kondisional terhadap kelas yang diberikan. Artinya, setiap fitur dianggap tidak mempengaruhi fitur lainnya setelah mempertimbangkan kelas.Untuk mengklasifikasikan data baru, algoritma menghitung probabilitas posterior untuk setiap kelas menggunakan teorema *Bayes* dan memilih kelas dengan probabilitas tertinggi.

Ada beberapa Jenis *Naïve Bayes*, Gaussian *Naïve Bayes* Digunakan ketika fitur-fitur dalam data diasumsikan mengikuti distribusi Gaussian (normal). Biasanya digunakan untuk data yang bersifat kontinu.Multinomial *Naïve Bayes* Cocok untuk data yang berupa count atau frekuensi, seperti dalam klasifikasi teks. Misalnya, menghitung berapa kali suatu kata muncul dalam dokumen.Bernoulli *Naïve Bayes* Digunakan ketika fitur adalah variabel biner (misalnya, hadir/tidak hadir). Sering digunakan dalam klasifikasi teks dengan representasi biner.

* 1. Kerangka Pemikiran

Berikut ini adalah kerangka pemikiran berdasarkan hasil identifikasi masalah dan dilakukan tahapan penyesuaian berdasarkan dasar teori yang telah diidentifikasi sebelumnya. Kerangka pemikiran penelitian sebagai berikut.



**Gambar 2.2.** Kerangka Pemikiran

Diagram alur seperti pada gambar 2.2. menggambarkan proses pengolahan data dan analisis sentimen dimulai dari aplikasi My JNE, yang berfungsi sebagai titik awal pengumpulan data. Data diperoleh melalui teknik scraping, yang mengacu pada ekstraksi informasi dari aplikasi My JNE dan situs web Google Play. Proses pengumpulan data ini memastikan bahwa informasi yang diperoleh adalah representatif dan relevan untuk analisis lebih lanjut.

Setelah data terkumpul, tahap preprocessing dimulai untuk mempersiapkan data agar siap dianalisis. Proses pertama dalam tahap ini adalah cleansing, di mana data dibersihkan dari karakter atau simbol yang tidak diperlukan, sehingga hanya teks yang bersih dan relevan yang dipertahankan. Selanjutnya, case folding diterapkan untuk mengubah seluruh teks menjadi huruf kecil, memastikan konsistensi dalam analisis dan menghindari perbedaan antara huruf kapital dan huruf kecil. Proses ini diikuti oleh tokenization, yang memecah teks menjadi unit-unit kecil seperti kata atau frasa, yang memungkinkan analisis yang lebih mendalam pada level kata.

Tahap berikutnya adalah normalisasi, yang mengubah kata-kata menjadi bentuk dasar atau standar mereka, mengurangi variasi dalam data teks. Filtering dilakukan untuk menghapus kata-kata yang tidak relevan atau terlalu umum, seperti stop words, yang tidak memberikan informasi tambahan dalam konteks analisis sentimen. Terakhir, stemming diterapkan untuk mengubah kata-kata menjadi bentuk dasar atau akar katanya, menyederhanakan teks dan memudahkan proses analisis lebih lanjut.

Setelah preprocessing, data siap untuk dianalisis menggunakan metode analisis yang telah dipilih. Di antara metode yang digunakan adalah K-Nearest Neighbors (K-NN), yang merupakan algoritma klasifikasi berbasis kedekatan data, dan Naive Bayes, yang merupakan algoritma probabilistik yang mengklasifikasikan data berdasarkan distribusi probabilitas. Metode ini memanfaatkan fitur-fitur yang telah diproses untuk menentukan kategori sentimen dari data.

Hasil akhir dari proses ini adalah pelabelan sentimen, di mana setiap data dikategorikan sebagai positif, negatif, atau netral berdasarkan analisis yang telah dilakukan. Diagram ini menggambarkan alur kerja umum dalam pemrosesan bahasa alami (NLP) untuk analisis sentimen, di mana data mentah melalui beberapa tahap pemrosesan dan analisis untuk menghasilkan informasi yang dapat digunakan untuk memahami sentimen dari data yang dianalisis.

* 1. Hipotesis

(jika ada)

1. METODE PENELITIAN
   1. Analisis Kebutuhan

Objek penelitian ini adalah salah satu jasa kurir terbesar diindonesia yaitu PT.Tiki jalur Nugraha Ekakurir atau yang biasa disebut JNE. adapun subjek Penelitian ini adalah Opini pengguna pengiriman JNE terhadap Pelayanan JNE melalui riview pada database *web*site google play yang hingga 17 Agustus 2024 telah mencapai sekitar 5 juta download lebih dengan 12 ribu lebih *review* sejak awal pertama kali aplikasi my jne dirilis dan akan terus bertambah jumlahnya seiring bertambahnya waktu.

* + 1. Populasi dan Sampel

Sampel dalam penelitian ini adalah semua data riview atau ulasan pengguna PT. TIKI Jalur Nugraha Ekakulir (JNE) dari database *web*site google play dengan aplikasinya yang bernama *My JNE* dengan total 12521 *review*. sedangkan sampel diambil dari suatu populasi yang mana pada penelitian ini dinggunakan sampel ulasan pengguna JNE sejak adanya *update* aplikasi *My JNE* versi terbaru yaitu 2.3.9, dari tanggal 30 Agustus 2016 Sampai dengan 17 Agustus 2024. dengan jumlah sampel 5.000 *review*.

* + 1. Metode Pengumpulan Data

Metode yang digunakan untuk proses pengumpulan data dari penelitian ini adalah sebagai berikut :

1. Web *Scraping*

Data pada penelitian ini adalah data sekunder yang diambil secara online dengan teknik *scraping* data menggunakan ekstensi Scraper yang ada di *Google* *Colab* dengan menggunakan *Python*. Melalui teknik scraper ini didapatkan informasi dari *web*site secara otomatis tanpa perlu menyalin secara manual.

1. Studi Pustaka

Studi pustaka dilakukan dengan menggunakan beberapa kajian literatur, buku, maupun referensi jurnal yang sekiranya berkaitan dengan tujuan, rumusan, batasan, dan metode penelitian.

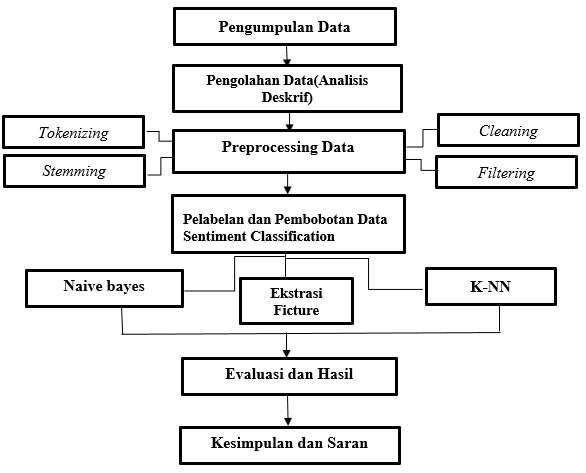
* + 1. Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian ini adalah data sekunder. Pada penelitian ini data sekunder berperan sebagai main data dalam proses pengolahan untuk mengetahui opini pengguna terhadap pelayanan yang diberikan oleh perusahaan JNE. Adapun sumber data yang digunakan dalam penelitian ini adalah *Dataset* *review*s terhadap kepuasan pengguna JNE terhadap pelayanan yang diberikan di *web*site *Google Play*, yang mana peneliti mengambil secara online dengan link berikut ini:

*https://play.google.com/store/apps/details?id=com.indivara.jneone&hl=id*.

Pengambilan data secara online ini dinamakan *web* *scraping* atau *scraping* data. Adapun *scraping* data pada penelitian ini dilakukan dengan menggunakan *Google* *Colab* dan *Python*. *Google* *Chrome* memiliki berbagai jenis ekstensi yang mana salah satunya adalah *Google* *Colab* yang berfungsi sebagai *web* scraper untuk mengambil data atau konten yang terdapat pada situs online secara gratis. *Google* *Colab* mengekstraksi atau mengambil data dari halaman *web* menggunakan bahasa pemrograman *Python* dan mengimpornya ke dalam *spreadsheet Microsoft Excel*. Data hasil ekstraksi ini kemudian akan diolah untuk dapat mengetahui opini pengguna JNE.

* + 1. Perancangan Penelitian



**Gambar 3.1.** Perancangan Penelitian

Berdasarkan diagram alir penelitian, langkah-langkah pada penelitian ini terdiri atas sebagai berikut:

1. Mengumpulkan Data

Pada tahap ini dilakukan pengumpulan data, namun berbeda dengan penelitian lainnya karena tidak dilakukan dengan datang langsung ke perusahaan. Pengumpulan data yang akan diolah dilakukan dengan mengekstraksi data *review* secara online dengan teknik *scraping* data menggunakan ekstensi dari *Google* *Chrome*, yaitu *Google* *Colab*. *Web* *scraping* dilakukan pada *review* pengguna JNE melalui aplikasi yang terdapat di *Google Play* yaitu *My JNE*.

1. Analisis Deskriptif

Analisis ini digunakan untuk menggambarkan ulasan atau opini pengguna JNE, waktu *review* pengguna, atau rating yang diberikan secara umum melalui aplikasi *My JNE* pada situs *web*site *Google Play*.

1. *Preprocessing Data*

*Preprocessing* berguna untuk menyeleksi data dan mengubahnya menjadi data yang terstruktur. Ini merupakan tahapan awal untuk mengubah struktur isi data menjadi format yang sesuai agar dapat diproses oleh algoritma *Naïve Bayes* *Classifier*. Proses *Preprocesing* ini dilakukan menggunakan *Software* R dan *Python*, dengan tahapan sebagai berikut:

* Cleaning: Tahap membersihkan dokumen dari kata yang tidak diperlukan, tanda baca (~!@#$%^&\*()\_-+=[]{}:;‟”.>,<), dan mengubah huruf dalam dokumen menjadi huruf kecil semua (*lowercase*).
* *Filtering*: Tahap menghapus kata-kata yang kurang bermakna atau tidak memiliki arti seperti kata: saya, dan, atau.
* *Tokenizing*: Tahap pemisahan teks menjadi potongan kalimat dan kata atau yang biasa disebut token.
* Stemming: Tahap pencarian kata dasar dari tiap kata pada teks.

1. Pelabelan dan Pembobotan Data

* Klasifikasi dengan *Naïve Bayes* *Classifier* dan *KNN*: Metode *Machine Learning* *Naïve Bayes* *Classifier* dan *KNN* digunakan untuk melakukan klasifikasi *review* menjadi kelas positif, negatif, maupun netral. Pada tahap ini, digunakan *Google* *Colab*.
* Visualisasi dan Asosiasi: Visualisasi dan asosiasi dilakukan pada semua data, baik yang berupa sentimen positif, negatif, maupun netral. Visualisasi ini dilakukan dengan mengekstraksi keseluruhan informasi sehingga dapat diketahui kata yang paling sering digunakan pada komentar pengguna. Bentuk dari visualisasi dan asosiasi ini nantinya akan berupa wordcloud dan barplot.

1. Evaluasi dan Hasil

Setelah mengetahui visualisasi dari sentimen negatif, langkah selanjutnya adalah membuat *Diagram Fishbone* untuk menganalisis faktor-faktor yang menjadi masalah di mata pengguna JNE serta mengidentifikasi hal yang perlu dilakukan untuk mengatasi permasalahan tersebut.

1. Penutup (Kesimpulan dan Saran)

Tahap terakhir dari penelitian ini adalah menarik kesimpulan dan memberikan saran berdasarkan hasil penelitian.

* + 1. Metode Analisis Data
  1. *Scraping Data*

Dalam penelitian ini, pengumpulan data dilakukan melalui proses *scraping* pada aplikasi *My JNE* yang tersedia di *Google Play*. Data yang diambil berupa ulasan pengguna yang diberikan antara tanggal 30 Agustus 2018 hingga 17 Agustus 2024. Teknik *scraping* ini memungkinkan peneliti untuk mengumpulkan informasi secara otomatis dari 5000 ulasan yang tersedia di halaman *Google Play*, memberikan dasar data yang diperlukan untuk analisis sentimen.

Proses *scraping* dilakukan dengan menggunakan skrip *Python* yang dirancang khusus untuk mengakses dan mengunduh data ulasan secara berkala selama periode penelitian. Data yang diperoleh kemudian disimpan dalam bentuk *Dataset* yang mencakup berbagai atribut seperti tanggal ulasan, rating, serta isi komentar pengguna. *Dataset* ini menjadi dasar bagi analisis lebih lanjut menggunakan metode *Naïve Bayes* *Classifier* dan *K-Nearest Neighbors* untuk mengidentifikasi pola sentimen dalam ulasan pengguna jasa pengiriman JNE.

Dalam proses *scraping*, data diambil secara bertahap untuk menghindari pemblokiran oleh *Google Play Store*. Untuk setiap permintaan data, script ini mengekstrak dan menyimpan data dalam bentuk format *CSV* yang mencakup semua atribut penting. Setiap baris dalam file *CSV* merepresentasikan satu ulasan pengguna, yang mencakup informasi seperti ID pengguna, rating yang diberikan, tanggal ulasan, dan teks ulasan itu sendiri. Proses ini memerlukan waktu yang cukup lama karena volume data yang besar, namun dengan pemrograman yang efisien, pengumpulan data dapat diselesaikan sesuai dengan target yang diharapkan.

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah pembersihan dan praproses data sebelum digunakan dalam analisis sentimen. Proses pembersihan meliputi penghilangan data duplikat, koreksi kesalahan penulisan, dan penghilangan karakter khusus yang tidak relevan. Selain itu, data ulasan yang terlalu pendek atau tidak memberikan informasi yang bermakna juga dihapus. Hasil akhir dari proses ini adalah *Dataset* yang bersih dan siap digunakan untuk tahap analisis dengan metode *Naïve Bayes* *Classifier* dan *K-Nearest Neighbors*. *Dataset* yang telah diproses ini memberikan gambaran yang lebih akurat mengenai sentimen pengguna terhadap layanan pengiriman JNE selama periode penelitian.

* 1. *Preprocessing*

Tahap selanjutnya setelah proses *scraping* adalah *Preprocesing*, yaitu langkah-langkah yang dilakukan untuk membersihkan dan mempersiapkan data agar siap digunakan dalam analisis sentimen. Pada tahap ini, data ulasan yang telah dikumpulkan akan melalui beberapa tahapan seperti normalisasi teks, penghilangan stopwords, dan tokenisasi. Normalisasi teks meliputi proses mengubah huruf menjadi *lowercase*, menghilangkan tanda baca, serta menghapus karakter atau kata yang tidak relevan. Stopwords, yaitu kata-kata umum yang tidak memiliki kontribusi signifikan terhadap analisis sentimen, dihapus untuk meningkatkan akurasi model. Setelah itu, teks ulasan dipecah menjadi token-token atau kata-kata yang akan digunakan sebagai input dalam model *Machine Learning*. Hasil dari *Preprocesing* ini adalah *Dataset* yang lebih bersih, terstruktur, dan siap untuk dianalisis menggunakan metode *Naïve Bayes* *Classifier* dan *K-Nearest Neighbors*, memungkinkan untuk menghasilkan wawasan yang lebih mendalam terkait sentimen pengguna terhadap layanan pengiriman JNE

* + 1. Model Akurasi Data *dengan Confusion Matrix*

Untuk menguji akura model algoritma Naïve Bayes dan K-Nearest Neighbors (KNN), dengan analisis berbasis Confusion Matrix untuk mengevaluasi performa model. Menurut Manning, Prabhakar, dan Hinrich (2009), Confusion Matrix adalah alat penting dalam machine learning yang menampilkan hasil klasifikasi ke dalam tabel yang mencerminkan prediksi benar dan salah. Matriks ini digunakan untuk menghitung akurasi, recall, dan presisi model.

* + 1. Akurasi (*Accuracy*)

Akurasi mengukur proporsi dari prediksi yang benar dibandingkan dengan total jumlah prediksi yang dibuat. Ini dihitung dengan membagi jumlah prediksi benar dengan jumlah total prediksi. Akurasi memberikan gambaran umum tentang seberapa sering model membuat prediksi yang benar. Akurasi dihitung dengan menjumlahkan semua prediksi yang benar (baik positif maupun negatif) dan membaginya dengan jumlah total prediksi yang dibuat.

(13)

* + 1. Presisi (*Precision*)

*Precision* menunjukkan proporsi dari prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dibuat. Dihitung dengan membagi jumlah prediksi benar yang positif dengan jumlah total prediksi positif. *Precision* penting untuk menilai keakuratan dari prediksi positif yang dihasilkan model. *Precision* dihitung dengan membagi jumlah prediksi benar yang positif (*True* *Positive*s) dengan jumlah total prediksi positif yang dibuat (jumlah *True* *Positive*s + False *Positive*s).

(15)

* + 1. *Recall* (Sensitivitas atau *True* *Positive* *Rate*)

*Recall* mengukur seberapa baik model dapat menangkap semua kasus positif yang benar-benar positif dalam *Dataset*. Dihitung dengan membagi jumlah prediksi benar yang positif dengan jumlah total kasus positif yang sebenarnya. *Recall* berguna untuk memahami seberapa efektif model dalam mengidentifikasi kasus positif yang langka. *Recall* dihitung dengan membagi jumlah prediksi benar yang positif (*True Positives*) dengan jumlah total kasus positif yang sebenarnya (jumlah *True Positives* + False Negatives).

(14)

* + 1. *F1 Score*

*F1-Score* adalah ukuran keseimbangan antara *Precision* dan *Recall*, memberikan rata-rata harmonik dari keduanya. Ini dihitung dengan mengalikan *Precision* dan *Recall*, kemudian membaginya dengan jumlah keduanya, dan mengalikan hasilnya dengan 2. *F1-Score* berguna ketika ada kebutuhan untuk menyeimbangkan *Precision* dan *Recall* dalam evaluasi model. *F1-Score* dihitung dengan mengalikan *Precision* dan *Recall*, kemudian mengalikannya dengan 2 dan membaginya dengan jumlah *Precision* dan *Recall*.

(16)

DAFTAR PUSTAKA

Afandi, I. R., & dkk. (2022). Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes. *Jurnal Linguistik Komputasional (JLK), 5*(2), 63-69.

Alita, D., & Rahman, A. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi, 8*(2), 50-58.

Cuatanto, R., & Sutomo, R. (Indonesian Journal of Computer Science). Comparative Analysis Between Naïve Bayes Algorithm and Decision Tree Loss Rate from Fire Disaster Data in DKI Jakarta Province. *Indonesian Journal of Computer Science, 12*(4), 1787-1796.

Ferryawan, R., Kusrini, & Wibowo, F. W. (2019). Analisis Sentimen Wisata Jawa Tengah Menggunakan Naϊve Bayes. *Jurnal INFORMA, 5*(3), 55-60.

Flores, V. A., Permatasari, P. A., & Jasa, L. (2020). Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword. *Majalah Ilmiah Teknologi Elektro, 19*(2), 157-162. doi:10.24843/MITE.2020.v19i02.P06

Giovani, A. P., Ardiansyah, Haryanti, T., Kurniawati, L., & Gata, W. (2020). ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI. *Jurnal TEKNOINFO, 14*(2), 116-124. doi:10.33365/jti.v14i2.679

Harahap, F., & dkk. (2021). Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat. *Jurnal Ilmiah Informatika, 9*(1).

Harun, A., & Ananda, D. P. (2021). Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decision Tree. *MALCOM: Indonesian Journal of Machine Learning and Computer Science, 1*(1), 58-63.

Irnawati, O., & Solecha, K. (2022). Analisis Sentimen Ulasan Aplikasi Flip Menggunakan Naïve Bayes dengan Seleksi Fitur PSO. *Jurnal Ilmiah Intech, 4*(2), 189-199.

Juniarsih, S., Ripantia, E. F., & Pratama, E. E. (2020). Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi. *JUSTIN (Jurnal Sistem dan Teknologi Informasi), 8*(3), 239-246. doi:10.26418/justin.v8i3.39118

Lestari, N. A., Akhriza, T. M., & Yuniar, E. (2020). METODE NAÏVE BAYES CLASSIFIER DENGAN TEXTBLOB UNTUK ANALISIS SENTIMEN TERHADAP PELAYANAN INDIHOME DAN FIRST MEDIA. *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK), 4*(1), 283-290.

Nishfi Ilmiah Huda, D., Prianto, C., & Awangga, R. M. (2023). Analisis Sentimen Perbandingan Layanan Jasa Pengiriman Kurir Pada Ulasan Play Store Menggunakan Metode Random Forest dan Decision Tree. *Jurnal Ilmiah Informatika, 11*(2), 151-155.

Nitami, M. T., & Februariyanti, H. (2022). Analisis Sentimen Ulasan Ekspedisi J&T Express Menggunakan Algoritma Naive Bayes. *Jurnal Manajemen Informatika & Sistem Informasi (MISI), 5*(1), 20-29. doi:10.36595/misi.v5i1

Pertiwi, M. W. (2019). Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, KNN dan SVM. *Inti Nusa Mandiri, 14*(1), 27-32.

Putri, D. S., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi Pospay dengan Algoritma Support Vector Machine. *Jurnal Ilmiah Informatika, 11*(1), 33-40.

Ramadhani, T., Sari, Y. A., & Santoso, E. (2021). Analisis Sentimen Masyarakat Indonesia terhadap Covid-19 pada Media Sosial Twitter menggunakan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 5*(12), 5680-5686.

Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *JURNAL INOVTEK POLBENG - SERI INFORMATIKA, 3*(1), 50-55.

Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *Jurnal IKRA-ITH Informatika, 3*(2), 1-8.

Romadoni, F., Umaidah, Y., & Sari, B. N. (2020). Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine. *Jurnal SISFOKOM (Sistem Informasi dan Komputer), 9*(2), 247-253. doi:10.32736/sisfokom.v9i2.903

Rosdiana, & dkk. (2019). Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar. *Proceeding SNTEI*, 87-93.

Samsir, & dkk. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA, 5*(1), 157-163. doi:10.30865/mib.v5i1.2604

Saraswati, M., & Riminarsih, D. (2020). Analisis Sentimen Terhadap Pelayanan KRL Commuterline Berdasarkan Data Twitter Menggunakan Algoritma Bernoulli Naive Bayes. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA), 25*(3), 225-230. doi:https://doi.org/10.35760/ik.2020.v25i3.3256

Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS, 10*(2), 681-686.

Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2020). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *Journal of Intelligent Systems and Computation*, 43-49.

Septiana, R. D., Susanto, A. B., & Tukiyat, T. (2021). Analisis Sentimen Vaksinasi Covid-19 Pada Twitter Menggunakan Naive Bayes Classifier Dengan Feature Selection Chi-Squared Statistic Dan Particle Swarm Optimization. *Jurnal Sistem Komputer dan Kecerdasan Buatan, 5*(1), 49-56.

Setiawan, H., Utami, E., & Sudarmawan, S. (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi dan Informatika), 5*(1), 43-51. doi:10.31603/komtika.v5i1.5189

Siti, & Mayary, J. (2020). Analisis Sentimen Pengguna Twitter terhadap Dompet Elektronik dengan Metode Lexicon Based dan K-Nearest Neighbor. *Jurnal Ilmiah Informatika Komputer, 25*(1), 1-12. doi:10.35760/ik.2020.v25i1.2411

Wahyudi, R., & Kusumawardhana, G. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *JURNAL INFORMATIKA, 8*(2), 200-207.

Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *JURNAL INFORMATIKA, 7*(1), 1-7.

Yasid, M., & Junaedi, L. (2019). Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naïve Bayes. *JURNAL ILMIAH INFORMATIKA, 7*(2), 83-91.